

# ENTERED



PCT

## RAW SEQUENCE LISTING

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:44

Input Set : N:\CrF4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

```

1 <110> APPLICANT: The Trustees of Columbia University in the City of New York
2 <120> TITLE OF INVENTION: Novel Tumor-Associated Marker
3 <130> FILE REFERENCE: 0575/60240-PCT
4 <140> CURRENT APPLICATION NUMBER: US/09/646,569A
5 <141> CURRENT FILING DATE: 2001-09-18
6 <160> NUMBER OF SEQ ID NOS: 28
7 <170> SOFTWARE: PatentIn version 3.1
9 <210> SEQ ID NO: 1
10 <211> LENGTH: 333
11 <212> TYPE: PRT
12 <213> ORGANISM: Human
13 <400> SEQUENCE: 1
14   Met Pro Leu Gly Leu Gly Arg Arg Lys Lys Ala Pro Pro Leu Val Glu
15   1           5           10           15
16   Asn Glu Glu Ala Glu Pro Gly Arg Gly Gly Leu Gly Val Gly Glu Pro
17           20           25           30
18   Gly Pro Leu Gly Gly Gly Gly Ser Gly Gly Pro Gln Met Gly Leu Pro
19           35           40           45
20   Pro Pro Pro Pro Ala Leu Arg Pro Arg Leu Val Phe His Thr Gln Leu
21           50           55           60
22   Ala His Gly Ser Pro Thr Gly Arg Ile Glu Gly Phe Thr Asn Val Lys
23           65           70           75           80
24   Glu Leu Tyr Gly Lys Ile Ala Glu Ala Phe Arg Leu Pro Thr Ala Glu
25           85           90           95
26   Val Met Phe Cys Thr Leu Asn Thr His Lys Val Asp Met Asp Lys Leu
27           100          105          110
28   Leu Gly Gly Gln Ile Gly Leu Glu Asp Phe Ile Phe Ala His Val Lys
29           115          120          125
30   Gly Gln Arg Lys Glu Val Glu Val Phe Lys Ser Glu Asp Ala Leu Gly
31           130          135          140
32   Leu Thr Ile Thr Asp Asn Gly Ala Gly Tyr Ala Phe Ile Lys Arg Ile
33           145          150          155          160
34   Lys Glu Gly Ser Val Ile Asp His Ile His Leu Ile Ser Val Gly Asp
35           165          170          175
36   Met Ile Glu Ala Ile Asn Gly Gln Ser Leu Leu Gly Cys Arg His Tyr
37           180          185          190
38   Glu Val Ala Arg Leu Leu Lys Glu Leu Pro Arg Gly Arg Thr Phe Thr
39           195          200          205
40   Leu Lys Leu Thr Glu Pro Arg Lys Ala Phe Asp Met Ile Ser Gln Arg
41           210          215          220
42   Ser Ala Gly Gly Arg Pro Gly Ser Gly Pro Gln Leu Gly Thr Gly Arg
43           225          230          235          240
44   Gly Thr Leu Arg Leu Arg Ser Arg Gly Pro Ala Thr Val Glu Asp Leu

```

## RAW SEQUENCE LISTING

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:44

Input Set : N:\Crf4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

```

45          245          250          255
46    Pro Ser Ala Phe Glu Glu Lys Ala Ile Glu Lys Val Asp Asp Leu Leu
47          260          265          270
48    Glu Ser Tyr Met Gly Ile Arg Asp Thr Glu Leu Ala Ala Thr Met Val
49          275          280          285
50    Glu Leu Gly Lys Asp Lys Arg Asn Pro Asp Glu Leu Ala Glu Ala Leu
51          290          295          300
52    Asp Glu Arg Leu Gly Asp Phe Ala Phe Pro Asp Glu Phe Val Phe Asp
53    305          310          315          320
54    Val Trp Gly Ala Ile Gly Asp Ala Lys Val Gly Arg Tyr
55          325          330
57 <210> SEQ ID NO: 2
58 <211> LENGTH: 720
59 <212> TYPE: DNA
60 <213> ORGANISM: Human
61 <400> SEQUENCE: 2
62    cacggggagg cggaggcagc ggcggcgggc ggcggcgggc cggcgggcggc ggagcagatc      60
63    ttctggtgac cccacttctc gctgctcatg ccgctgggac tggggcgccg gaaaaaggcg      120
64    cccctctag  tggaaaatga ggaggctgag ccaggccgtg gagggctggg cgtgggggag      180
65    ccagggcctt tgggcggagg tgggtcgggg ggccccaaa tgggcttgcc cccctcccc      240
66    ccagccctgc ggccccgct tgtgtccac accagctgg cccatggcag tccactggc      300
67    cgcacgagg ggttcaccaa cgtcaaggag ctgtatggca agattgccga ggccttcgc      360
68    ctgccactg ccgaggtgat gttttgcacc ctgaacaccc acaaagtga catggacaag      420
69    ctctggggg gccaaatcgg gctggaggac ttcatcttcg cccacgtgaa ggggcagcgc      480
70    aaggaggtgg aggtgttcaa gtcggaggat gcactcgggc tcaccatcac ggacaacggg      540
71    gctggctacg ccttcatcaa gcgcacaaag gagggcagcg tgatcgacca catccacctc      600
72    atcagcgtgg gcgacatgat cgaggccatt aacgggcaga gcctgctggg ctgccggcac      660
73    tacgaagtgg cccggtgct caaggaactg ccccgaggcc gtaccttcac gctgaagctc      720
75 <210> SEQ ID NO: 3
76 <211> LENGTH: 9
77 <212> TYPE: PRT
78 <213> ORGANISM: Human
79 <400> SEQUENCE: 3
80    Lys Leu Leu Gly Gly Gln Ile Gly Leu
81    1          5
83 <210> SEQ ID NO: 4
84 <211> LENGTH: 10
85 <212> TYPE: PRT
86 <213> ORGANISM: HUMAN
87 <400> SEQUENCE: 4
88    Ser Leu Leu Gly Cys Arg His Tyr Glu Val
89    1          5          10
91 <210> SEQ ID NO: 5
92 <211> LENGTH: 6263
93 <212> TYPE: DNA
94 <213> ORGANISM: Human
95 <400> SEQUENCE: 5
96    catcagcggg cgggggtgtc gccgaacagg ctgctccgca gagcccgccg cgaccccgcg      60
97    ccgccccgcc ccgcggcctg cctgccagag gagccgaggg ggccgcccct cgcccaacct      120

```

## RAW SEQUENCE LISTING

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:44

Input Set : N:\Crf4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

98	gcccgcacatg ggggaaccccc ggcccaggcg tgcctggtcac catgacaaca gagacaggcc	180
99	ccgactctga ggtgaagaaa gctcaggagg aggccccgca gcagcccagag gctgctgccg	240
100	ctgtgaccac cctgtgacc cctgcaggcc acggccaccc agaggccaac tccaatgaga	300
101	agcatccatc ccagcaggac acgcggcctg ctgaacagag cctagacatg gaggagaagg	360
102	actacagtga ggccgatggc ctttcggaga ggaccacgcc cagcaaggcc cagaaatcgc	420
103	cccagaagat tgccaagaaa tacaagagtg ccatctgccg ggtcactctg cttgatgcct	480
104	cggagtatga gtgtgagggtg gagaaacatg gccggggcca ggtgctgttt gacctggtct	540
105	gtgaacacct caacctccta gagaaggact acttcggcct gaccttctgt gatgctgaca	600
106	gccagaagaa ctggctggac cctccaagg agatcaagaa gcagatccgg agtagccctt	660
107	ggaattttgc cttcacagtc aagttctacc cgcctgatcc tgcccagctg acagaagaca	720
108	tcacaagata ctacctgtgc ctgcagctgc gggcagacat catcacgggc cggctgccat	780
109	gctcctttgt cacgcattgc ctactgggct cctacgctgt gcaggetgag ctgggtgact	840
110	atgatgctga ggagcatgtg ggcaactatg tcagcgagct ccgcttcgcc cctaaccaga	900
111	cccgggagct ggaggagagg atcatggagc tgcataagac atataggggg atgaccccg	960
112	gagaagcaga aatccacttc ttagagaatg ccaagaagct ttccatgtac ggagtagacc	1020
113	tgcaccatgc caaggactct gagggcatcg acatcatgtt aggcgtttgt gccaatggcc	1080
114	tgctcatcta ccgggaccgg ctgagaatca accgctttgc ctggcccaag atcctcaaga	1140
115	tctcctacaa gaggagtaac ttctatatca agatccggcc tggggagtat gagcaatttg	1200
116	agagcacaat tggctttaag ctcccaaac accggtcagc caagagactg tggaggtct	1260
117	gcctcgagca tcatacatc ttccggctgg tgtccctga gccccaccc aagggtcttc	1320
118	tgggtgatgg ctccaagtcc cggtagctg ggaggacca ggcacagact cggcaggcca	1380
119	gcgcctcat tgaccggcct gcaccttct ttgagcgttc ttccagcaaa cggtagacca	1440
120	tgtcccgag ccttgatgga gcagagttct cccgccagc ctcggtcagc gagaacctg	1500
121	atgcaggggc tgacgggtgac aagcgggatg aggatggcga gtctggggg caacggtcag	1560
122	aggctgagga gggagaggtc aggaactcaa ccaagatcaa ggagctaaag ccggagcagg	1620
123	aaaccacgcc gagacacaag caggagttct tagacaagcc agaagatgtc ttgctgaagc	1680
124	accaggccag catcaatgag ctcaaaagga ccctgaagga gcccaacagc aaactcatcc	1740
125	accgggatcg agactgggaa cgggagcgca ggtgccctc ctccccgcc tccccctccc	1800
126	ccaagggcac ccctgagaaa gccaatgaga gaggagggt gagggagggc tccgaggaga	1860
127	aagtcaaac accacgtccc cgggccccag agagtgcac aggcgatgag gaccaggacc	1920
128	aggagaggga caggtgttc ctgaaggaca accacctggc cattgagcgc aagtgtcca	1980
129	gcatacggg cagctctacg tctagctgg aggtgaggt ggacttcacg gtcattggtg	2040
130	actaccatgg cagcgcttc cagacttct cccgagcct gcctgagctc gaccgggaca	2100
131	aaagcgactc ggacactgag ggctgtgt tctccggga tctcaacaag ggggccccca	2160
132	gccaggatga tgagtctggg ggcattgagg acagcccgga tcgaggggccc tgctccaccc	2220
133	cggatatgcc ccagtttgag cccgtgaaaa cagaaacct gactgtcagc agtctggcca	2280
134	ttagaaagaa gattgagcgg gaggcgtac tgcagaccag agtctccgct atggataaca	2340
135	cccagcaggt tgatgggagt gcctcagtgg ggaggaggt catagcaacc actccctcca	2400
136	tcaccacgga gaccatatcg accaccatgg agaacagtct caagtccggg aagggggcag	2460
137	ctgccatgat ccaggccca cagacgggtg ccacggaaat ccgttctctt tctccgatca	2520
138	tcgggaaaga tgtcctcacc agcacctacg gcgccactgc ggaaaccctc tcaacctcca	2580
139	ccaccaccca tgtcaccaaa actgtgaaag gagggttttc tgagacaagg atcgagaagc	2640
140	gaatcatcat tactggggat gaagatgtcg atcaagacca ggccctggct ttggccatca	2700
141	aggaggccaa actgcagcat cctgatatgc tggtaaccaa agctgtcgta tacagagaaa	2760
142	cagaccatc cccagaggag agggacaaga agccacagga atcctgacct ctgtgaagag	2820
143	atcctggcat ttctgggtcca acccaagcca gagaaccatt aagaaggggc cttcattctg	2880
144	gattctccga cgcaacactg acgtcccagc tgcgacgtac tgcactgat aggaaacacc	2940
145	gaagggaaaa gcatatatat atagatatat agagatatag atatatatac gggaaacacc	3000
146	gcataccttgc actgctgctg gggctggcag agcagttggc tgacagcaac aaccgacatc	3060

## RAW SEQUENCE LISTING

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:44

Input Set : N:\Cr4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

147	tgaacaccta	catttccctt	gcagacaaat	tgaagaactg	gtgggatttt	tttcaagaaa	3120
148	aaaaattata	taataactat	aatcccttgc	tcaccccttt	ccccgcgcaa	ataagaaacg	3180
149	caagccagac	cacgatgatt	gtagaagtcc	ctcccgccct	ggttctgcac	gttacagtta	3240
150	gcagacgagc	aattccattt	gttctttctc	agcatctcta	aggcccaact	gaatgcaaag	3300
151	gaaaacactt	gcacagcaaa	gcaagagaag	tcacagcagc	aagacacgca	cagtcaacca	3360
152	ttttccgaga	aaaaaagaaa	attccccact	tggaaagaaa	gaggaggaac	actggattct	3420
153	tactttctgg	atcttgacac	tgggtgcgaa	aacctacctt	cctctctccc	gcctccccct	3480
154	acctcaact	ctcaatgtct	tgtgtctcatt	ttctgtctcg	gtccctctct	cccccttccc	3540
155	ccttccccca	ccccacaccc	ttcaccctct	gtgtccctgt	ccttctgagg	gccactgcag	3600
156	atgactctcc	tttgaaatga	gaaaaagaaa	agaaagcaag	aacagaaaaac	gaagccacag	3660
157	gaagggaagt	agacattgta	tgtttatggt	ttctcattat	gaagggtgcag	cttgtaggag	3720
158	gtttgtacgg	atgtgctttg	aagttatgta	tattacatat	aacaggaaaa	aatattaata	3780
159	aacagtgtctg	gtaagtatga	agctgacatt	ctaaaattat	aattatctga	ctgtgattga	3840
160	tgtatccctga	ggttccctaga	tctcaactgaa	ctggcccagc	taaggagacc	tggactctgg	3900
161	gtgtgggttg	gtcacagta	ggggctgacg	ggttcagtgt	agtaatactg	tgtgtggtgt	3960
162	ttgtaattgg	ttgattgggtg	gggaggggtg	gggggccccta	atggagaggt	gtgggtttgg	4020
163	caagaaagaa	gcaacacaga	tgtcgtcccc	aaaatgccag	ttcaagacac	cttctccctg	4080
164	ccccctgggt	agtaacagtc	agggcctggt	ctgtgctcag	gtactgggtc	ccagtctggg	4140
165	actctgctgc	tgaagttgcc	acagtagagg	tccttggtct	agtccttctc	tcctacggg	4200
166	gcttgccttg	gttttcagtc	ttctctctct	ttctctcttt	tttttttttt	tgccacattc	4260
167	tgcccttccc	tgacccctatt	gtaataacca	actccatctc	caaagggagg	tgggtgctct	4320
168	agccattgta	gaagatgggtg	gctttaacct	gactgtctaa	aaattcccag	ctaagccttt	4380
169	tcctctactc	tcttccctgt	tctgaatcat	ttcttctctt	caggccaaag	tagccatggt	4440
170	aaggaggctt	catggggcag	acctgaaag	atcaaaactg	catttgcaaa	gcctccccct	4500
171	gtcccaggac	aaagctgaga	ctgacgggtg	atgttgctca	taggctccag	ctctgcataa	4560
172	gaccttggt	tggagacctc	cctctcagtc	aacagctgaa	ctctgagctt	gtgccagaa	4620
173	attaccccaa	gaccacagga	accttccaag	aagctcccat	cacaagcttg	gcattgctct	4680
174	ctgccacaag	tgggttccct	caggcttgct	tgccacaagc	tacttctctg	agctcagaaa	4740
175	gtgccccttg	atgagggaaa	atgtcccact	gcactgcgaa	tttctcagtt	ccattttacc	4800
176	tcccagtcct	ccttctaaac	cagttaataa	attcattcca	caagtattta	ctgattacct	4860
177	gcttgtgcca	gggactattc	tcaggctgaa	gaagggtgga	ggggaggggc	gaacctgagg	4920
178	agccacctga	gccagcttta	tatttcaacc	atggctggcc	catctgagag	catctcccca	4980
179	ctctcgccaa	cctatcgggg	catagccag	ggatgcccc	aggcgcccca	ggttagatgc	5040
180	gtccccttgg	cttgtcagtg	atgacataca	ccttagctgc	ttagctggtg	ctggcctgag	5100
181	gcagggcagg	aaatcagaat	agcatttget	tctctgggca	aatgggaagt	tcagcggggc	5160
182	agcagaatca	gtggcattcc	ccctggtgca	ggccgggtgg	tcactccaa	ctccccctga	5220
183	gtgtagcagc	acactttcca	tacaccaggt	tctttctaca	atcctggtgg	aaaagccaca	5280
184	gaaccttctt	cctgcccctc	ttgagagttc	cccccttttc	tgggtcaaga	gctggagtgg	5340
185	tggctccatc	ctctctgggc	cacttcggtc	taggaactca	tctttgcagg	aaccaggagt	5400
186	cctgagcaca	ctgaacacac	ctcagaggga	ggatccttgt	tgtggatttt	gcacctggct	5460
187	ttggggcagg	ggtgaagtga	ccaggcttag	cttgtggagt	ttatgggcca	ccagggtttg	5520
188	gggaaatcac	catcccgcg	atgctgtgac	ctcccttcta	cggagatgca	ggcagtgcc	5580
189	cgaggaggga	ggggacctgc	aaagctagaa	tctagggcac	tgtttccctc	ccatccctct	5640
190	ctttgtagag	aatagagacg	tttgtcttgt	ctgtcttcaa	cctacttttc	cttttctctt	5700
191	ttttgtttct	catcctctct	gtgccacctc	tcacccag	aggccatgta	gcatagtgga	5760
192	aaaagtcctt	gagggcggtt	aggagttctg	ggtgaccatc	ctggctcagc	tcctaactca	5820
193	ccatgtgaca	tcaggctatc	cccattcccc	ctcttgggcc	tcagtttccc	gacttgcaaa	5880
194	ataagcagaa	agaaccagat	gctctccagg	gtctttttct	actttgtctat	ctcatgggtc	5940
195	ttcattttct	cttattttgt	tttctctgga	tcttttccat	ctgagggtac	aggaagtacc	6000

## RAW SEQUENCE LISTING

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:44

Input Set : N:\Crf4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

```

196   aggacctgtt tcagtttttg aatcctgcaa gcacattcca agactggcct gaaactgcat   6060
197   gagcaacatc actcgaaata attttttttt tcaaaagcac cttacaacc aattgcatg   6120
198   ctgtcctgtt cttttttact cacacccttc tctcctttct cgcccccatg cccccccacc   6180
199   tcagtgtctc gtgctgtatg cgtgtgctct ctgttcttgt atactcaata taagtgaat   6240
200   aaatgtgttt gatgctgaac cat                                     6263
202 <210> SEQ ID NO: 6
203 <211> LENGTH: 933
204 <212> TYPE: PRT
205 <213> ORGANISM: Human
206 <400> SEQUENCE: 6
207   Ser Ala Gly Gly Gly Val Ala Glu Gln Ala Ala Pro Gln Ser Pro Pro
208   1           5           10           15
209   Arg Pro Arg Ala Ala Pro Pro Arg Gly Leu Pro Ala Arg Gly Ala Glu
210           20           25           30
211   Gly Ala Ala Pro Arg Pro Thr Cys Pro Thr Trp Gly Thr Pro Gly Pro
212           35           40           45
213   Gly Val Leu Val Thr Met Thr Thr Glu Thr Gly Pro Asp Ser Glu Val
214           50           55           60
215   Lys Lys Ala Gln Glu Glu Ala Pro Gln Gln Pro Glu Ala Ala Ala Ala
216   65           70           75           80
217   Val Thr Thr Pro Val Thr Pro Ala Gly His Gly His Pro Glu Ala Asn
218           85           90           95
219   Ser Asn Glu Lys His Pro Ser Gln Asp Thr Arg Pro Ala Glu Gln Ser
220           100          105          110
221   Leu Asp Met Glu Glu Lys Asp Tyr Ser Glu Ala Asp Gly Leu Ser Glu
222           115          120          125
223   Arg Thr Thr Pro Ser Lys Ala Gln Lys Ser Pro Gln Lys Ile Ala Lys
224           130          135          140
225   Lys Tyr Lys Ser Ala Ile Cys Arg Val Thr Leu Leu Asp Ala Ser Glu
226   145          150          155          160
227   Tyr Glu Cys Glu Val Glu Lys His Gly Arg Gly Gln Val Leu Phe Asp
228           165          170          175
229   Leu Val Cys Glu His Leu Asn Leu Leu Glu Lys Asp Tyr Phe Gly Leu
230           180          185          190
231   Thr Phe Cys Asp Ala Asp Ser Gln Lys Asn Trp Leu Asp Pro Ser Lys
232           195          200          205
233   Glu Ile Lys Lys Gln Ile Arg Ser Ser Pro Trp Asn Phe Ala Phe Thr
234           210          215          220
235   Val Lys Phe Tyr Pro Pro Asp Pro Ala Gln Leu Thr Glu Asp Ile Thr
236   225          230          235          240
237   Arg Tyr Tyr Leu Cys Leu Gln Leu Arg Ala Asp Ile Ile Thr Gly Arg
238           245          250          255
239   Leu Pro Cys Ser Phe Val Thr His Ala Leu Leu Gly Ser Tyr Ala Val
240           260          265          270
241   Gln Ala Glu Leu Gly Asp Tyr Asp Ala Glu Glu His Val Gly Asn Tyr
242           275          280          285
243   Val Ser Glu Leu Arg Phe Ala Pro Asn Gln Thr Arg Glu Leu Glu Glu
244           290          295          300
245   Arg Ile Met Glu Leu His Lys Thr Tyr Arg Gly Met Thr Pro Gly Glu

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 03/27/2003  
PATENT APPLICATION: US/09/646,569A      TIME: 13:16:45

Input Set : N:\Crf4\03272003\I646569.raw  
Output Set: N:\CRF4\03272003\I646569A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:15; N Pos. 23,42,48,1105

Seq#:17; N Pos. 23,42

**VERIFICATION SUMMARY**

DATE: 03/27/2003

PATENT APPLICATION: US/09/646,569A

TIME: 13:16:45

Input Set : N:\Crf4\03272003\I646569.raw

Output Set: N:\CRF4\03272003\I646569A.raw

L:885 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15 after pos.:0

M:341 Repeated in SeqNo=15

L:953 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17 after pos.:0